**Computational modelling**

Following previous recommendations (Wilson & Collins, 2019) computational modeling was performed in two steps. First, we compared the relative fit of several alternative models (see *Model comparison*) and secondly, we examined whether the winning model could recover parameters from simulated data (see *Parameter recovery*).

**Models**

**Model 1**: Random responding, which assumes that participants push buttons at random with some bias for one option over the other. The reason for including the model is to detect participants who do not engage in the task. The bias is captured with parameter *b* (ranging between 0 -1), hence the probability of choosing option 1 and option 2 is:

and

For two bandits, this model has only one free parameter, *b*.

**Model 2**: Noisy win-stay-lose-shift, is a heuristic for making choices that repeats rewarded actions and switches away from the rewarded action with a probability of *ε.*

The probability of choosing option k is:

(1)

Where *ct* = 1, 2 is the choice at trial *t*, and *rt* = 0,1 the reward at trial *t*. The model has one free parameter, that is, the probability of switching from the rewarded action, *ε.*

**Model 3**: The Q-learning model (Watkins & Dayan, 1992) is a variant of Rescorla-Wagner model that take each history of previous outcomes into account to learn the long-term value of choosing an action. In the Q-learning model, the value of option *k*, is updated in response to the value of the reward *rt* according to:

(2)

where *α* is the learning rate and captures the extent to which the participant takes the difference between the expected and received outcome termed the *prediction error* into account, updating the value on the next trial. When *α* is 0, the value of the chosen option is not updated, and no learning takes place. When *α* is 1, the value of is fully updated according to the most recent outcome.

We used the Softmax choice rule (Equation 3) to compute the probabilities for each choice. The choice rule assumes that the participant choses the choice with the highest value (exploiting), but occasionally make explorative choices by choosing the low value option.

(3)

is the inverse temperature, controlling the degree of exploration (mistakes), ranging from  *= 0* for complete random choices and  *=∞* for completely deterministic choices. That is, an agent with a low value of will often deviate from the learned contingencies by choosing the action previously associated with lower values (explore). This may be adaptive if the environment is volatile or if the task is not fully known. In contrast, an agent with a high will chose the option associated with the highest learned value with a high probability, and only rarely explore.

**Model 4:** Noisy win-stay-lose-shift rule with separate exploration rates flowing rewards and losses, that is, it has two free parameters: *ε+* and *ε-*.

**Model 5**: Q-learning model with separate learning rates, *α+* and *α-,* otherwise it is identical to Model 3 and hence it has three free parameters.

**Model comparisons**

The different models were compared using the Bayesian information criterion (BIC). As can be seen in *Table S1*, Model 1 (Q-learning with a single learning rate *α*) resulted in the lowest mean BIC values in both groups and conditions.

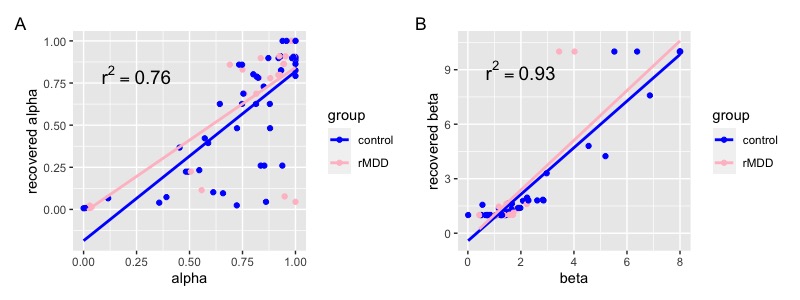
**Table S1.** *Mean BIC Values for the Compared Models. The Mean BIC Values are Bolded for the Winning Model, Q-Learning Model with one α Value.*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Model* | *Description* | *Number of free parameters* | *Mean BIC*  *(face)* | | | *Mean BIC*  *(non-social)* | | |
|  |  |  | Full sample | HC | rMDD | Full sample | HC | rMDD |
| 1 | Q learning (1 *α*) | 2 | **72.93** | **66.62** | **75.72** | **72.93** | **71.70** | **75.12** |
| 2 | Q learning (2 *α*) | 3 | 75.58 | 69.44 | 78.67 | 75.58 | 74.54 | 77.44 |
| 3 | Noisy WSLS (1 *ε*) | 1 | 80.37 | 78.42 | 84.38 | 80.37 | 79.98 | 81.06 |
| 4 | Noisy WSLS (2 *ε*) | 2 | 75.39 | 71.16 | 77.22 | 75.39 | 74.18 | 77.55 |
| 5 | Random responding + side bias | 1 | 100.90 | 99.40 | 97.08 | 100.90 | 100.42 | 101.76 |

*Note.* BIC = Bayesian Information Criterion. WSLS = Win-Stay-Lose-Shift. The lowest BIC values for each condition and group are marked in bold.

**Parameter Recovery**

We simulated data based on the original parameter values from 780 agents, which generated new data from which we estimated new parameters. We took the mean values from the recovered parameters and plotted them against the original parameter values. Results show at relatively high correlations, however, as shown in Figure S1 there is not a linear relationship between original and recovered parameters. Together with the non-significant results in the statistical analysis, our conclusion is that even though the Q-learning model provides a relatively good fit, it does not capture what the participants do. The interpretation is supported by the fact that the proportions of correct choices differ significantly, and hence a correct model should be able to capture this difference.



*Figure S1.* Correlation (*r2*) between original and recovered parameters from a model including, learning rate (alpha) and exploration (beta). Panel A: Correlation between original alpha values (x-axis) and recovered alpha values (y-axis) for control (blue line) and rMDD (pink line). Panel B: Correlation between original beta values (x-axis) and recovered beta values (y-axis) for control (blue line) and rMDD (pink line).

**Eye-tracking data pre-processing**

A pupil dilation response was calculated for each trial and defined as the mean proportional change in pupil size during 1) the expectation phase (1.5 seconds), and 2) the feedback phase (2 seconds) relative to baseline pupil size. Baseline pupil size was measured during one second directly preceding the stimulus onset (i.e., while the fixation cross was shown). The time window of the baseline interval was defined based on visual inspection of the overall pupil curve. Visual fixations were identified using an I-VT filter with velocity threshold set to 30 degrees of the visual field. All trials with less than 50% valid pupil samples were discarded (1.47% of trials during the expectation phase, and 4.39% during the feedback phase respectively).

Preliminary analyses were conducted using linear mixed effects models (see *Statistical analyses*). These analyses showed no relation between response times and pupil dilation during the expectation phase (*b* = 0.13, *SE* = 0.12, *t =* 1.07, *p* = .28), or the feedback phase (*b =* 0.07, *SE* = 0.12, *p* =.58). Additional analyses were conducted to test whether eye movements during the experiment were related to pupil dilation. For each recording, we calculated correlations between (filtered) pupil size and the Euclidian distance from the centre of the screen. This analysis showed only a weak average negative correlation (mean *r* = -.07, *SD* = 0.11). As can be seen in *Table 1,* no group differences were found in the average number of fixations during the experiment.